Cycles in random k-ary maps and the poor performance of random random number generation

Robin Pemantle [1,2]

ABSTRACT:

Knuth [Knu97] shows that iterations of a random function perform poorly on average as a random number generator. He proposes a generalization in which the next value depends on two or more previous values. This note demonstrates, via an analysis of the cycle length of a random k-ary map, the equally poor performance of a random instance in Knuth's more general model.

Keywords: birthday problem, poisson approximation, iterated functions

Subject classification: Primary: 65C10

[1]Research supported in part by National Science Foundation grant # DMS 0103635

[2]The Ohio State University, Department of Mathematics, 231 W. 18th Avenue, Columbus, OH 43210, pemantle@math.ohio-state.edu

1 Introduction

1.1 Statement of problem

In the introduction to his second volume, Knuth [Knu97] discusses the computer generation of pseudo-random numbers. He gives several cautionary tales about poor methods of generating these, including a function whose description is so complicated that it mimics iterations of a function chosen at random from all functions from $\{1, \ldots, 10^{10}\}$ to itself. The exercises (see Exercises 11–15 on page 8 of [Knu97]) then lead one through an analysis of a model where a function from $[m] := \{1, \ldots, m\}$ to itself is chosen uniformly at random. The poor performance of this pseudo-random number sequence is related to the cycle structure of a random map and is well understood. In particular, one may see readily that the average length of the cycle of numbers produced from a random seed is of order $\sqrt{m}$ and the cycle length from the best seed is not much longer.

Knuth then proposes the following generalization [Knu97, Problem 19, page 9, labeled M48]. A function f is chosen uniformly from among the $m^{(m^k)}$ functions from $[m]^k$ to $[m]$. Given an initial vector of values in $[m]^k$ for $(X_1, \ldots, X_k)$, an infinite sequence of values is produced by the rule

$$X_{n+k} = f(X_n, \ldots, X_{n+k-1})\,. \tag{1.1}$$

The problem is to determine the average length of the period of this eventually periodic sequence if the initial k seeds are chosen at random, and to answer as well some related questions: what is the chance that the eventual period has length 1, what is the average maximum cycle length over all seeds, what is the chance that there is no seed giving a cycle of length 1, and what is the average number of distinct eventual cycles as the seed varies?

1.2 Heuristic

A thumbnail computation shows that one might expect equally poor performance from this multiple dependence model. Let $W_n \in [m]^k$ denote $(X_n, \ldots, X_{n+k-1})$ and let $\mu < \tau$

be such that $W_{\tau-k+1} = W_{\mu-k+1}$ but the values of W up to $W_{\tau-k}$ are distinct; thus the eventual period is $\tau - \mu$ and the length of the sequence of X values before repeating is $\tau - k$. Although the values $\{W_n : n \geq 0\}$ are no longer independent in the generalized model, one may hope that they are nearly independent, so that the value of the random quantity τ is well approximated by the number N of IID uniform draws from a population of size $M := m^k$ needed to obtain the first repeated value. This is the classical "birthday problem" (see Example (3d) on page 33 and the discussion on page 49 of [Fel50]). It is known that the mean of N in the birthday problem is of order $\sqrt{M}$ and more precisely that $N^2/2M$ converges in distribution to a mean-one exponential as $M \to \infty$. One would therefore expect (Theorem 1.2 below) that $\tau^2/2m^{k/2}$ converges to a mean-one exponential as well.

1.3 Background

The problem of random number generation is of coure fundamental to the theory of computing. There are many classes of problems, including various factoring, counting and optimization problems, for which randomized algorithms give solutions much faster (on average) then any known non-randomized algorithm. The use of randomization in practice, is if anything, more widespread than would be justified by theoretical results. Monte Carlo methods are ubiquitous in the areas of scientific computing, for example, and banks use vast tables of pre-generated random numbers to price derivative securities.

The need for sources of effectively random numbers has mushroomed with the explosion in computational ability. Meanwhile, just as in the 1950's, the best sources are pseudo-random number generators, which are, for all their potential flaws, less prone to misbehave than are physical sources of randomness [Nie92].

To construct objects with random or chaotic properties, it is often easiest to pick them at random. Expander graphs, for example, are easy to construct at random but difficult to construct deterministically. When it comes to pseudo-random number generation, it is particularly appealing to use random or generic generators. The problem of finding a class

of generators, most of which generate good pseudo-random sequences, is therefore one of great interest to theorists and practitioners alike.

The random unary map is a natural starting point for exploration of random pseudo-random sequences, but also underlies many other phenomena in probability theory, ranging from random trees to Brownian paths. For this reason it has been studied in great detail (see [Kol86, AB82, FO90, AP94]) often with specific attention to the time before repetition, as in [AP94]. Its short cycle time, hence unsuitability for random number generation, has long been understood, as is evident from the discussion in all editions of [Knu97]. This has led to a huge industry in random number generation (see, e.g., [Nie92]). Classes of functions on $[m]$ or $[m]^k$ such as linear feedback registers or congruential sequences are sought whose periods are much longer than periods of functions chosen at random. Interestingly, the short cycle length can be a boon rather than always a drawback. Pollard's rho-algorithm [Pol75] relies on the $\sqrt{m}$ cycle length to find any prime factor p of n in time $O(\sqrt{p})$; Pollard's heuristic argument is in fact borne out empirically (see the discussion in [SF96, pages 466–471]).

Researchers studying random number generation via k-ary maps appear to have taken for granted that, as in the unary case, these random maps cycle in a relatively short time. Settling this, however, was a problem stated already in the 1981 edition of [Knu97], given a rating of [M48], and left unsolved.

1.4 Results

The purpose of this note is to show that the thumbnail computations are correct. All of the questions posed in Knuth may be correctly answered using the independence heuristic.

In the case $k = 2$ the following result holds. Let m be given and let τ, as defined above, be the least value for which $W_{\tau-1}$ repeats a previous value W_j for $j < \tau - 1$.

Theorem 1.1 *As $m \to \infty$, the quantity $\frac{\tau^2}{2m^2}$ converges in distribution to an exponential of mean 1. Furthermore, all moments of $\frac{\tau^2}{2m^2}$ converge to moments of the exponential. In*

particular,

$$\mathbb{E}\,\tau \;\sim\; m\sqrt{\frac{\pi}{2}}\,;$$
$$\mathrm{Var}(\tau) \;\sim\; \left(2-\frac{\pi}{2}\right)m^2\,.$$

For general k we will show:

Theorem 1.2 *For any fixed k and x, as $m \to \infty$,*

$$\mathbb{P}\left(\frac{\tau^2}{2m^k} \geq x\right) \to \exp(-x)\,.$$

Despite the fact that the arguments are straightforward, a careful analysis may be justified for several reasons. First, the question has gone unanswered for sufficiently long (and not for lack of interest) that the methods of analysis, though straightforward, must not be readily apparent. It will therefore be useful to introduce to the computer science community two techniques that are well known to probabilists and theoretical statisticians. In order to illustrate the range of available techniques for this kind of analysis, two different proofs will be presented.

The first is a direct, combinatorial analysis and will be presented for the case $k = 2$ (as is stated in Problem 16 to be the first interesting generalization), though it can easily be generalized to larger k. It relies on the concept of *hazard rate*, well known in actuarial circles. A brief introduction to this concept is given in the next section. The second analysis uses the Poisson approximation machinery of [AGG89], which relies on some technical lemmas of [BE83] and concepts developed by Chen and Stein in the 1980's. This method is discussed in Section 3. Although the Chen-Stein method is not elementary, the present application of this machinery is straightforward.

A second reason for undertaking this analysis is re-inject questions about basic random models into the stream of scientific discussion. Often a result on a basic model will rekindle interest in simple variations and lead to a branch of research previously overlooked by a community that tends towards depth-first research agendas.

Lastly, and perhaps most importantly, understanding the behavior of iterations of a random k-ary map may prove useful for other random models. Just as short cycles of random unary maps have been used in Pollard's rho-algorithm and elsewhere, it is not hard to imagine the random k-ary map underlying various other structures, algorithms and heuristics.

2 Time before repetition when $k = 2$

Given any sequence $\mathbf{X} := \{X_1, X_2, \ldots\}$ of values in $[m]$, we define the positive integer $\tau = \tau(\mathbf{X})$ as above to be minimal so that $W_{\tau-k+1} = W_{\mu-k+1}$ for some $k \leq \mu < \tau$ (where W_j are sub-words of length k of the $\mathbf{X}$ vector, as in the introduction). When the $\mathbf{X}$ vector is random, we let $\mathcal{F}_n$ denote the σ-field $\sigma(X_1, \ldots, X_n)$. Compare the distributions of the vector $(\tau, X_1, \ldots, X_\tau)$ under two different measures for $\mathbf{X}$: (a) when $\mathbf{X}$ satisfies the recursion (1.1) with $X_0, \ldots, X_{k-1}$ IID uniform on $[m]$ and (b) when $\mathbf{X}$ is an IID sequence of uniform draws from $[m]$. Under both (a) and (b), the conditional probability of $X_{n+1} = j$ given $\mathcal{F}_n$ is $1/m$ as long as $\tau > n$. The vector $(\tau, X_1, \ldots, X_\tau)$ therefore has the same distribution under either law on $\mathbf{X}$. The main subject of our analysis it the distribution of τ and other quantities measurable with respect to $\mathcal{F}_\tau$. We will therefore assume throughout that $\mathbf{X}$ is an infinite IID uniform sequence.

Let m be given. In the remainder of this section, the dependence length, k, is fixed at two; thus τ is minimal so that $(X_{\tau-1}, X_\tau) = (X_{\mu-1}, X_\mu)$ for some $\mu < \tau$. Arguments will be based on the hazard rate principle, an informal statement of this is the following.

> The exponential lieftime of mean 1 is charactarized by the property that at any given time t, if you haven't died yet, then you have chance dt of dying in the interval $(t, t+dt)$. Now suppose we observe an individual and determine at every time t a chance $f(t)\,dt$ for him to die in the next dt. This will be by defintion a random variable measureable with respect to events up to time t (weight, smoking habits, and so on). It is no longer an easy matter to determine the unconditional lifetime distribution of the population

from typical information about the random function f, but we can say somthing under a random time change. If the lifetime always has a conditional future density, then the amount of hazard up to the time of death, defined by $\int_0^T f(t)\,dt$ where T is the death time, will always be exponentially distributed with mean 1 (see [Jeu80, Prop. 3.28]). This principle is often used for bounding the tails $\mathbb{P}(T > x)$, e.g., by $e^{-c} + \mathbb{P}(f(x) < c)$.

The proof of Theorem 1.1 is an elementary chain of asymptotic equalities. Letting $\mathcal{E}(1)$ denote an exponential of mean 1, we will show that

$$\mathcal{E}(1) \stackrel{\mathcal{D}}{=} \text{ hazard } \approx \text{ linearized hazard } \approx \frac{\tau^2}{2m^2} \, .$$

The first equality states that in the present discrete-time context, one may still find a time change under which the lifetime is an exponential. The trick is to introduce auxilliary randomness to form the fractional part of the lifetime. This is Lemma 2.1 below, which is true for any stopping time. The remaining asymptotic inequalities then deal with approximations introduced by the discrete time steps and by a small amount of unpredictability in the hazard rate (cf. the methods of [Pem96]).

Lemma 2.1 *Let $\tau > 0$ be a stopping time on a probability space $(\Omega, \mathbb{P})$ with respect to a filtration $\{\mathcal{F}_n : n \geq 0\}$ and let $h(k)$ be the random variable defined by $h(k) = -\log(1 - A_k)$ on the event that $\tau > k$ and arbitrarily otherwise, where*

$$A_k := \mathbb{P}(\tau = k + 1 \,|\, \mathcal{F}_k) \, .$$

Let

$$h = h^* + \sum_{j=0}^{\tau-2} h(j)$$

where h^* is a random variable, whose conditional distribution given $h(\tau - 1) = x$ is a mean 1 exponential conditioned to be less than x. Suppose $\sum h(j) = \infty$ almost surely. Then h is distributed exactly as a mean 1 exponential.

Proof: Given $x > 0$, let G_n be the event that $\sum_{j=0}^{n-2} h(j) < x \leq \sum_{j=0}^{n-1} h(j)$. Then $G_n \in \mathcal{F}_{n-1}$ and

$$\begin{aligned}
\mathbb{P}(h \geq x) &= \sum_{n=1}^{\infty} \mathbb{P}(h \geq x, G_n) \\
&= \sum_{n=1}^{\infty} \mathbb{E}\mathbb{P}(h \geq x, G_n \,|\, \mathcal{F}_{n-1}) \\
&= \sum_{n=1}^{\infty} \mathbb{E}\mathbf{1}_{G_n} \mathbb{P}(h \geq x \,|\, \mathcal{F}_{n-1}) \\
&= \sum_{n=1}^{\infty} \mathbb{E}\mathbf{1}_{G_n} \left(\prod_{j=0}^{n-2} e^{-h(j)} \right) e^{-(x - \sum_{j=0}^{n-2} h(j))} \\
&= \sum_{n=1}^{\infty} \mathbb{E}\mathbf{1}_{G_n} e^{-x} \\
&= e^{-x}
\end{aligned}$$

since $\sum_j h(j) = \infty$ implies $\sum_n \mathbf{1}_{G_n} = 1$. □

The remainder of the proof of Theorem 1.1 involves combinatorial specification of the hazard rate. Apply the hazard rate lemma to the quantity τ in the statement of Theorem 1.1, resulting in quantities $h(k)$ and h satisfying $h \stackrel{\mathcal{D}}{=} \mathcal{E}(1)$. Define Y_k to be the number of $j < k$ for which $X_j = X_k$. An easy lemma is:

Lemma 2.2 *As $m \to \infty$, $m^{-1/2} \max\{Y_k : k \leq \tau\} \to 0$ in probability.*

Proof: Keep a tally of how many times each value has been seen in the sequence $X_1, X_2, \ldots$. Since these are independent draws, it is evident that with probability $O(e^{-c_\epsilon \sqrt{m}})$ for some $c_\epsilon > 0$, no value is taken on $\epsilon\sqrt{m}$ times before every value is taken on $\epsilon\sqrt{m}/2$ times. At a time T_ϵ when every value has been taken on $\epsilon\sqrt{m}/2$ times, the hazard function h is at least $c(\epsilon)m$, where $c(\epsilon)$ is a constant not depending on m. It follows for fixed $\epsilon > 0$ that the probability of $\tau > T_\epsilon$ is exponentially small in m, and consequently that the probability of $\max\{Y_k : k \leq \tau\}$ exceeding $\epsilon\sqrt{m}$ is at most the sum of two probabilities that are exponentially small in $\sqrt{m}$, and hence that it tends to zero as $m \to \infty$. □

Recast the definition of A_k in terms of Y_k,

$$A_k = \mathbb{P}(\tau = k+1 \,|\, \mathcal{F}_k) = \frac{Y_k}{m}\,,$$

to obtain the following immediate consequence.

Corollary 2.3 *For every $\epsilon > 0$ there is a $c_\epsilon > 0$ such that*

$$\left| h - \sum_{j=0}^{\tau-1} h(j) \right| \leq \epsilon m^{-1/2}$$

with probability at least $1 - c_\epsilon^{-1}(\exp c_\epsilon \sqrt{m})$.

Proof: By the definition of h,

$$0 \geq h - \sum_{j=0}^{\tau-1} h(j) \geq -h(\tau-1) = \log(1 - \frac{Y_{\tau-1}}{m})\,. \tag{2.1}$$

By Lemma 2.2 this is at most $\epsilon m^{-1/2}$ except on a set of measure tending to zero exponentially in $\sqrt{m}$ for each fixed ϵ. $\square$

The cumulative linearized hazard rate

$$H(k) := \sum_{j=1}^{k} A_j$$

is close to $\sum_{j=0}^{k} h(j)$ but easier to work with. We will see that

$$\sum_{j=0}^{k} h(j) \approx H_k \approx \frac{\binom{k}{2}}{m^2}\,.$$

To quantify the last approximation, for $j \leq m$, let $T_k(j)$ be the number of $i \leq k$ for which $X_i = j$. Then, counting pairs of occurrences of each value, an alternate definition of $H(k)$ is:

$$H(k) = \sum_{j=1}^{m} \frac{\binom{T_k(j)}{2}}{m}\,.$$

Lemma 2.4 *If $k^2/m \to \infty$, then*

$$\frac{H_k}{\binom{k}{2}/m^2} \to 1$$

in probability.

PROOF: Denote the first moment, second moment and variance of H_k by μ_k, S_k and V_k respectively. We may compute these as follows. $\mu_k = E\binom{T_k(1)}{2}$. We compute μ_k as the expected number of pairs (i,j) of indices at most k for which $X_i = X_j = 1$. Clearly then

$$\mu_k = \frac{\binom{k}{2}}{m^2}\,.$$

Compute $m^2 S_k$ as $mET_k(1)^2 + m(m-1)ET_k(1)T_k(2)$. Counting ordered pairs of unordered pairs for which $X_u = X_v = 1$ and $X_w = X_x = 2$, we see that

$$ET_k(1)T_k(2) = \frac{\binom{k}{2}\binom{k-2}{2}}{m^4}\,.$$

In a similar way, allowing for (w,x) to have two, one or zero elements in common with (u,v), we get that

$$ET_k(1)^2 = \frac{\binom{k}{2}}{m^2} + \frac{\binom{k}{2}2(k-2)}{m^3} + \frac{\binom{k}{2}\binom{k-2}{2}}{m^4}\,.$$

Summing gives

$$S_k = \binom{k}{2}\binom{k-2}{2}(\frac{m(m-1)}{m^6} + \frac{m}{m^6}) + \frac{\binom{k}{2}2(k-2)}{m^4} + \frac{\binom{k}{2}}{m^3}.$$

Then

$$V_k = S_k - \mu_k^2 = \frac{\binom{k}{2}}{m^3} - \frac{\binom{k}{2}}{m^4}$$

and

$$\frac{V_k}{\mu_k^2} = \frac{m-1}{\binom{k}{2}}\,.$$

The lemma now follows from Chebyshev's inequality. $\square$

Remark: In order to prove convergence of all moments, one must estimate $\mathbb{E}H(k)^p$ for integers $p > 2$. There is an expansion analogous to the equation $m^2 S_k = m\mathbb{E}T_k(1)^2 +$

$m(m-1)\mathbb{E}T_k(1)T_k(2)$. Say that a descending vector of positive integers is a partition of m if $\lambda = (\lambda_1, \ldots, \lambda_{\#\lambda})$ and $\sum_{j=1}^{\#\lambda} \lambda_j = m$. Let T_k^λ denote the product $\prod_{j=1}^{\#\lambda} T_k(j)^{\lambda_j}$. Then

$$m^p \mathbb{E}H(k)^p = \sum_\lambda m^{\#\lambda}(1+O(m^{-1}))\mathbb{E}T_k^\lambda$$

where the sum runs over partitions of m. Here, the multiplier $m^{\#\lambda}(1+O(m^{-1}))$ gives the number of ways of choosing distinct $j_1, \ldots, j_{\#\lambda}$. When λ is the partition $(1, \ldots, 1)$, the leading term of the sum is

$$m^p \frac{\prod_{j=0}^{p-1} \binom{k-2j}{2}}{m^{2p}}$$

leading to a contribution of $(1+O(m^{-1}))(k^2/(2m^2))^p$. For any other λ, $\mathbb{E}T_k^\lambda$ is a sum of terms of the form $O(k/m)^a$ with $1 \le a \le p$. Each of these terms appears in $\mathbb{E}H(k)^p$ with the multiplier $m^{\#\lambda - p}$, which is $O(m^{-1})$. The total number of these terms is bounded, so it follows that when $k/m > \epsilon$,

$$\mathbb{E}H(k)^p = (1+O(m^{-1}))\left(\frac{k^2}{2m^2}\right)^p . \tag{2.2}$$

In other words, for $k/m \ge \epsilon$, $2m^2 H_k/k^2$ converges to 1 in each L^p as $m \to \infty$, uniformly in k.

Proof of Theorem 1.1: Convergence in distribution will follow from a comparison of $H(k)$ and $\sum_{j=0}^k h(j)$. From the definitions,

$$\begin{aligned}
\sum_{j=0}^{k\wedge\tau-1} h(j) &= \sum_{j=1}^{k\wedge(\tau-1)} -\log(1-A_k) \\
&= \sum_{j=1}^{k\wedge(\tau-1)} A_k + O(A_k)^2 \\
&= H(k\wedge(\tau-1))\left(1+O\left(\max_{j\le k\wedge(\tau-1)} A_j\right)\right) \\
&= H(k\wedge(\tau-1))\left(1+O\left(\max_{j\le k\wedge(\tau-1)} \frac{Y_j}{m}\right)\right) .
\end{aligned} \tag{2.3}$$

By Lemma 2.2, this shows that $\sum_{j=1}^{\tau-1} h(j)/H(\tau-1) \to 1$ in probability as $m \to \infty$. Since

$\tau^2/m \to \infty$ in probability as $m \to \infty$, Lemma 2.4 may be applied to show that

$$\frac{2n^2}{\tau^2} \sum_{j=0}^{\tau-1} h(j) \to 1 \tag{2.4}$$

in probability as $m \to \infty$. Together with Corollary 2.3, this implies that

$$\frac{2m^2}{\tau^2} h \to 1$$

in probability as $m \to \infty$, and convergence in distribution of $\tau^2/(2m^2)$ to $\mathcal{E}(1)$ then follows from the hazard rate lemma.

To extend this to convergence of higher integral moments, argue as follows. We know that

$$\mathcal{E}(1) \stackrel{\mathcal{D}}{=} h\,. \tag{2.5}$$

Let $||\cdot||_p$ denote the L^p norm. From Corollary 2.3 we see that

$$||\frac{h}{\sum_{j=0}^{\tau-1} h(j)}||_p \to 1 \tag{2.6}$$

as $m \to \infty$. Let G be the event that $\max\{Y_k : k < \tau\}$ is at most $m^{-1/2}$. It was shown in the proof of Lemma 2.2 that the probability of G^c decays exponentially in $\sqrt{m}$. It was already shown in (2.3) that

$$\left|\frac{\sum_{j=0}^{\tau-1} h(j)}{H(\tau-1)}\right| = 1 + O(m^{-1/2})$$

on G, which, together with the decay of $\mathbb{P}(G^c)$ faster than any polynomial, leads to

$$||\frac{\sum_{j=0}^{\tau-1} h(j)}{H(\tau-1)}||_p \to 1 \tag{2.7}$$

as $m \to \infty$. Finally, the estimate (2.2) in the case $k^2/m \geq \epsilon$ together with convergence of τ^2/m to ∞ in probability and monotonicity of $H(k)$ in k imply that

$$||\frac{H(\tau-1)}{\tau^2/(2m^2)}||_p \to 1 \tag{2.8}$$

as $m \to \infty$. The chain (2.5)–(2.8) of asymptotic equivalences in L^p proves the last statement of the theorem. $\square$

3 Analysis of $X_{n+k} = f(X_n, \ldots, X_{n+k-1})$ for any $k \geq 2$ via Poisson approximation

In this section we will prove Theorem 1.2. Let $N := \lfloor \sqrt{2m^k x} \rfloor$. Then $\tau^2/(2m^k) \geq x$ if and only if the values of W_n for $0 \leq n \leq N$ are distinct. Let Z be the number of pairs (i, j) for which $0 \leq i < j \leq N$ and Theorem 1.2 is an immediate consequence of:

Lemma 3.1 *The total variation distance between the law of Z and a Poisson of mean x is $o(1)$ as $m \to \infty$.*

The proof of Z is via the Chen-Stein Poisson Approximation method. The classical Poisson approximation result [Dur96, page 137] says that if events are

- independent,
- each has probability at most ϵ, which is going to zero,
- and the sum of the probaiblities converges to λ,

then the number that occur converges in law to a Poisson of mean λ. The Chen-Stein method is a means of weakening the independence assumption. It suffices that most events be independent, and the rest not too dependent. Stein [Ste88] developed an abstract framework for quantifying such limit laws. Later authors such as [AGG89] provided useful hypotheses for getting numerical bounds on the distance, in total variation, to a Poisson distribution.

Proof of Lemma 3.1: For the duration of this proof, α and α' will be shorthand for (i, j) and (i', j') respectively. Let S denote the set of α for which $0 \leq i < j \leq N$. Let G_α denote the event that $W_i = W_j$. Define $p_\alpha = \mathbb{P}(G_\alpha)$ and $p_{\alpha\alpha'} = \mathbb{P}(G_\alpha \cap G_{\alpha'})$.

Let $B(\alpha)$ be the set of α' for which $|y - y'| < k$ for some $y \in \{i, j\}$ and $y' \in \{i', j'\}$. Note that for $\alpha' \notin B(\alpha)$, the event $G_{\alpha'}$ that $W_{i'} = W_{j'}$ is measurable with respect to $\{X_s : |s - i|, |s - j| \geq k\}$. Therefore,

$$G_\alpha \text{ is independent of } \sigma(G_{\alpha'} : \alpha' \notin B(\alpha)) \,. \tag{3.1}$$

Define

$$
\begin{aligned}
b_1 &:= \sum_\alpha \sum_{\alpha' \in B(\alpha)} p_\alpha p_{\alpha'} \,; && (3.2)\\
b_2 &:= \sum_\alpha \sum_{\alpha \neq \alpha' \in B(\alpha)} p_{\alpha\alpha'} \,. && (3.3)
\end{aligned}
$$

The quantities b_1 and b_2 are quantities appearing under the same name in [AGG89, Theorem 1]; their quantity b_3 is zero due to the independence relation (3.1). The conclusion of [AGG89, Theorem 1] is that $|\mathbb{P}(Z=0) - \exp(\mathbb{E}Z)| < b_1 + b_2$. It remains to identify $\mathbb{E}Z$ and to bound b_1 and b_2 from above.

Observe first the claim that for any α, $p_\alpha = m^{-k}$. This is obvious for $|i-j| \geq k$. But in fact for any i and j, G_α occurs if and only if $X_{i+r} = X_{j+r}$ for $0 \leq r < k$. For $j = i+s$ with $0 < s < k$, the values of $X_i, \ldots, X_{i+s-1}$ may be chosen arbitrarily, and there will be precisely one set of values of $X_{i+s}, \ldots, X_{i+s+k-1}$ for which G_α occurs, proving the claim. It follows that

$$\lambda := \mathbb{E}Z = \sum_\alpha p_\alpha = m^{-k}\binom{N}{2} = (1+o(1))x \,. \tag{3.4}$$

Observe next that the cardinality of $B(\alpha)$ is at most $8kN$, since the number of pairs (i', j') with i' within k of i is at most $2kN$, and similarly for the other three possibilities. It follows immediately that

$$b_1 \leq (\sum_\alpha p_\alpha)(8kN)m^{-k} \leq (8\sqrt{2} + o(1))kx^{3/2}m^{-k/2} \,. \tag{3.5}$$

Finally, we bound b_2 from above. Let $B_0(\alpha)$ denote the set of α' for which both of i' and j' are within k of either i or j. Then $|B_0(\alpha)| < (4k)^2$.

Claim: for $\alpha \neq \alpha' \in B_0(\alpha)$,

$$p_{\alpha,\alpha'} \leq m^{-k-1} \,.$$

Assume without loss of generality that $j < j'$, since the other cases, $j > j'$, $i < i'$ and $i > i'$ are similar. Then

$$p_{\alpha,\alpha'} \leq p_\alpha \mathbb{P}(G_{\alpha'} \,|\, X_n : n < j') \leq m^{-k}m^{-1} \,,$$

proving the claim.

For $\alpha' \notin B_0(\alpha)$, one conditions on $\{X_s : |s-i| < k \text{ or } |s-j| < k\}$ to see that $p_{\alpha\alpha'} = m^{-2k}$. One then has

$$\begin{aligned} b_2 &= \sum_{\alpha} \left[\sum_{\alpha' \in B_0(\alpha)} p_{\alpha\alpha'} + \sum_{\alpha} \sum_{\alpha' \in B_0(\alpha)} p_{\alpha\alpha'} \right] \\ &\leq \sum_{\alpha} \left[16k^2 m^{-k-1} + (8kN) m^{-2k} \right] \\ &\leq 8k^2 N^2 m^{-k-1} + 4kN^{3/2} m^{-2k} \\ &= (1+o(1))(16k^2 \lambda m^{-1} + (42^{3/4} \lambda^{3/2} k m^{-k}) \end{aligned} \tag{3.6}$$

as $m \to \infty$. Combining (3.4) - (3.6) establishes that $b_1 + b_2 = o(1)$ and $\mathbb{E}Z - x = o(1)$, which completes the proof of Theorem 1.2. □

4 Further discussion

Let U and $\mathcal{E}(1)$ be independent with U uniform on $[0,1]$ and $\mathcal{E}(1)$ exponential of mean 1. The following extension of the distributional convergence results may be proved. Recall that μ is the index for which $X_\mu, \ldots, X_{\tau-1}$ is the first full period of the eventually periodic sequence of pseudo-random numbers.

Theorem 4.1 *As $m \to \infty$, the pair (μ, τ) converges in distribution to $(U\mathcal{E}(1), \mathcal{E}(1))$.*

Complete proof of the extensions in this section will not be given, but the argument, along the lines of the first analysis, is as follows. Fix an integer r and break the hazard rate for the occurrence of τ into r components. The j^{th} component at time n is the hazard rate for the occurrence of $\tau = n+1$ and $(j-1)/r \leq \mu < j/r$. A lemma analogous to Lemma 2.4 shows that the r hazards accumulate at asymptotically equal rates, and a lemma analogous to the hazard rate lemma then shows the asymptotic uniform distribution of μ/τ over the r bins given $\lfloor r\,\tau \rfloor$. Sending r to infinity completes the argument.

An analysis of the probability of landing in a cycle of length 1 is easiest along the lines of the Poisson approximation. Indeed, the number of occurrences of W_n of the form $(j, \ldots, j)$ for some $j \leq m$ by time k is well approximated by a Poisson of mean km^{1-k}; the number of these followed by one more j is then nearly a Poisson of mean km^{-k}. Since τ is of order $m^{k/2}$, one sees that the mean number of these occurrences by time τ is $\Theta(m^{-k/2})$, so this gives the order of magnitude of the chance of being caught in a cycle of length 1. On the other hand, the probability that some seed results in a cycle of length 1 is the chance that one of the m words $(j, \ldots, j)$ maps to itself, which rapidly approaches $1 - e^{-1}$ as $m \to \infty$.

An upper bound on the maximum value of τ over all seeds is obtained as follows. In the spirit of Theorems 1.1 and 1.2, the probability that $\tau^2 > 2(1+\epsilon)m^k(k \log m)$ can be shown to be close to $\exp(-(1+\epsilon)k \log m)$. Indeed, while Theorems 1.1 and 1.2, as written, compute $\mathbb{P}(\tau^2 > (2m^k)x)$ only when x is fixed, the arguments are sufficient to handle poly-logarithmic growth of x, that is $x \leq (\log m)^p$. Specifically, the four chains in the asymptotic equalities when x grows at this rate are: the exact equality $h \stackrel{\mathcal{D}}{=} \mathcal{E}(1)$ as before; the difference between h and $\sum_{j-0}^{\tau-1} h(j)$ small in every L^p; the linearization error in replacing h by H changes the likelihood of exceeding a hazard of x from e^{-x} to $e^{-x+o(x)}$, and the ratio between $H(k)$ and its deterministic counterpart $k^2/(2m^2)$ is small as long as x is not too small (as before). One may then extend the estimate to slowly growing x:

$$\mathbb{P}(\tau^2 > 2(1+\epsilon)km^k \log m) \sim m^{-(1+\epsilon)k} \, .$$

Since there are m^k seeds, this gives

$$\mathbb{P}\left[\tau^* > \sqrt{b\, k\, m^k\, \log m}\right] \to 0 \tag{4.7}$$

for any $b > 2$, where τ^* is the supremum over seeds of the value of τ for a fixed random f.

A interesting theoretical problem, perhaps do-able with only a modest amount of effort, would be to make this more precise and give a sharper estimate. For iterations of a unary map, the distribution of τ^* is known. Its mean was shown in [FO90, Theorem 7] to be asymptotic to an explicit contant multiple of $\sqrt{m}$. Later the scaling limit of $m^{-1/2}\tau^*$ was shown to exist [AP94] and an explicit formula given [AP02, Theorem 1].

Problem: *Let $\tau^*(m,k)$ be the supremum over all seeds of the value of τ for iterations of one random function $f : [m]^k \to [m]$. Show that $m^{-k/2}\tau^*$ converges weakly to the same limit as described in [AP94, AP02].*

If this proves difficult, perhaps it could at least be shown that $\tau^* = O(m^{k/2})$ in probability, that is, that the extra factor of log under the radical in (4.7) is superfluous.

Acknowledgement: The author would like to thank Philippe Flajolet for bringing this problem to light at the 8^{th} Analysis of Algorithms meeting in Strobl.

References

[AP94] Aldous, D. and Pitman, J. (1994). Brownian bridge asymptotics for random mappings. *Rand. Struct. Algor.* **5**, 487–512.

[AP02] Aldous, D. and Pitman, J. (2002). The asymptotic distribution of the diameter of a random mapping. *C. R. Math. Acad. Sci. Paris* **334**, 1021–1024.

[AB82] Arney, J. and Bender, E. (1982) Random mappings with constraints on coalescence and number of origins. *Pacific Journal of Mathematics* **103**, 269–294.

[AGG89] Arratia, R., Goldstein, L. and Gordon, L. (1989). Two moments suffice for Poisson approximation: the Chen-Stein method. *Ann. Probab.* **17**, 9–25.

[BE83] Barbour, A. and Eagleson, G. (1983). Poisson approximation for some statistics based on exchangeable trials. *Adv. Appl. Prob.* **15**, 585–600.

[Dur96] Durret, R. (1996). *Probability: Theory and Examples.* Duxbury Press: Belmont, CA.

[Fel50] Feller, W. (1950). *An introduction to probability theory and its applications, vol. 1.* John Wiley & Sons: New York.

[FO90] Flajolet, P. and Odlyzko, A. M. (1990). Random mapping statistics. In *Advances in Cryptology*, J.-J. Quisquater and J. Vandewalle, Eds., vol. 434 of *Lecture Notes in Computer Science*, Springer Verlag, pp. 329–354.

[Jeu80] Jeulin, T. (1980). *Semi-martingales et grossissement d'une filtration.* Lecture Notes in Mathematics No. 833. Springer: New York.

[Knu97] Knuth, D. (1997). *The art of computer programming, vol. 2: semi-numerical algorithms. Third edition.* Addison-Wesley: Reading, MA.

[Kol86] Kolchin, V. F. *Random Mappings*. Optimization Software Inc., New York, 1986. Translated from *Slučajnye Otobraženija*, Nauka: Moscow.

[Nie92] Niederreiter, H. (1992). *Random Number Generation and Quasi–Monte–Carlo Methods*. No. 63 in CBMS–NSF Regional Conference Series. Society for Industrial and Applied Mathematics: Philadelphia.

[Pem96] Pemantle, R. (1996). Maximum variation of total risk. *Stat. Prob. Letters* **28**, 285–289.

[Pol75] Pollard, J. M. (1975). A Monte Carlo method for factorization. *BIT* **15**, 331–334.

[SF96] Sedgewick, R. and Flajolet, P. (1996). *An Introduction to the Analysis of Algorithms.* Addison-Wesley Publishing Company: New York.

[Ste88] Stein, C. (1986). *Approximate computation of expectations.* No. 7 in the Institute of Mathematical Statistics Lecture Notes-Monograph Series. IMS: Hayward, CA.